EXHIBIT L

OMNIBUS BROWN DECLARATION

A Course in Econometrics

Arthur S. Goldberger

Harvard University Press
Cambridge, Massachusetts
London, England
1991

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper, and its binding materials have been chosen for strength and durability.

Library of Congress Cataloging-in-Publication Data

Goldberger, Arthur Stanley, 1930–
A course in econometrics / Arthur S. Goldberger.
p. cm.
Includes bibliographical references and index.
ISBN 0–674–17544–1 (alk. paper)
1. Econometrics. I. Title.
HB139.G634 1991
330'.01'5195—dc20 . 90–42284
CIP

—which is the only region where power is wanted in the present situation: see Exercise 22.4.

We have learned, as an equivalent to the t-test, the F-test that uses the statistic $v_j^\circ = (b_j - \beta_j^\circ)^2/\hat{\sigma}_{b_j}^2$, rejecting the null if $v_j^\circ > d$ where $G_1(d) = 0.95$, with $G_1(\cdot)$ being the cdf of the $F(1, n - k)$ distribution. The two approaches are equivalent because $v_j^\circ = (u_j^\circ)^2$ and $d = c^2$. But the F-statistic $v_j^\circ = (u_j^\circ)^2$ disregards the sign of $b_j - \beta_j^\circ$, so it is not attractive for use when the alternative is one-sided.

For a joint hypothesis with one-sided alternatives, no t-test is available. The F-statistic

$$v^\circ = (\mathbf{t} - \boldsymbol{\theta}^\circ)'\mathbf{D}(\mathbf{t} - \boldsymbol{\theta}^\circ)/(p\hat{\sigma}^2),$$

treats positive and negative misses symmetrically, so it is not attractive for tests against one-sided alternatives. For a discussion of appropriate procedures, see Gouriéroux et al. (1982) and Wolak (1987).

22.4. Choice of Significance Level

Suppose that you are asked to test the null hypothesis $\beta_j = 0$ against the alternative $\beta_j \neq 0$, in a sample with $n - k = 120$. You obtain the test statistic $u_j^\circ = 1.82$. Critical values from the $\mathcal{N}(0, 1)$ table are $c = 1.96$ at the 5% level and $c = 1.64$ at the 10% level. With $1.64 < 1.82 < 1.96$, the null would be accepted at the 5% level, but rejected at the 10% level. The same piece of evidence that will accept $\beta_j = 0$ at the 5% level will reject it at the 10% level. The interval between 1.64 and 1.96 is a "zone of opportunity." Indeed, whatever numerical value the sample delivers, a diligent researcher can force acceptance by setting the significance level low enough (e.g., 1% or 0.5%) or can force rejection by setting the significance level high enough (e.g., 10% or 20%).

How *should* a researcher choose the significance level? Econometrics texts offer little, if any, guidance. In statistics texts, the discussion focuses on the power of the test—the probability of rejecting the null hypothesis as a function of the true parameter value.

Generally power declines as the significance level declines: see Exercise 22.4. Moving from the 5% to the 1% significance level not only reduces the probability of rejecting a true null, but also reduces the probability of rejecting a false null. The first reduction is desirable, the second is undesirable.

There is a trade-off. To resolve the trade-off, statistics texts recommend a cost-benefit calculation: if the net cost of accepting a false null is less than the net cost of rejecting a true null, then choose a low significance level. Although this cost-benefit approach should be congenial to economists, the 5% level is almost always used in the empirical economics literature. It is hardly plausible that distinct cost-benefit calculations underlie that ubiquitous level. Occasionally, the 10% and 1% levels are used. Reading closely, you may well be able to spot the occasions on which those levels replace 5%. If an author really wants to accept the null, she may switch to the 1% level; if an author really wants to reject the null, he may switch to the 10% level. When such switches do not suffice, you may see such language as "barely significant at the 1% level" (a hint that the author really wants to accept) or "almost significant at the 10% level" (a hint that the author really wants to reject).

This state of affairs may seem very unsatisfactory, but the textbook recommendation of a cost-benefit calculation is not appealing either. For academic research reports, neither the costs nor the benefits of the test decision are clear. It is rare for an economic agent to undertake real-world action upon reading a test outcome reported in a journal article. At most what may happen is that readers' beliefs shift in the light of the evidence. So, in almost all applied economic contexts, the significance level is necessarily a matter of convention rather than of calculation.

It follows that readers should not take an author's announcement of significance or nonsignificance as authoritative. Regardless of the author's choice of significance level and announcement of a decision, sensible readers will have to decide for themselves whether the evidence is weighty or fragile. Regardless of how the author phrases the test decision, the burden remains on readers to assess whether the sample evidence against the null (the magnitude of the test statistic) is strong enough to induce a change in their beliefs.

A couple of lessons for writers emerge:

• It is usually bad practice to say "significant [or nonsignificant] at the 5% level," without reporting the magnitude of the test statistic. (It is even worse practice to announce "significance" or "nonsignificance" without specifying a null hypothesis. In particular, the zero null may not be the interesting null.)

• A useful alternative to the test statistic is a report of its "*P*-value," or "marginal significance level," which is the level at which the observed

test statistic would be just significant. For example, suppose that a $\chi^2(p)$ test is conducted, the cdf being $G_p(\cdot)$. If $w°$ is the observed test statistic, then its *P*-value is $\alpha° = 1 - G_p(w°)$. The null would be rejected at all significance levels higher than $\alpha°$, and accepted at all significance levels lower than $\alpha°$. So the *P*-value gives readers more information than is contained in the binary report "accept" or "reject."

22.5. Statistical versus Economic Significance

A strong case can be made that hypothesis testing is widely abused in empirical economics: see McCloskey (1985). In many research reports, the author's conclusions emphasize the statistical significance, rather than the economic significance, of the coefficient estimates. Yet, a coefficient estimate may be "very significantly different from unity" (by the *t*-test), while that difference is economically trivial. Or the difference may be "not significantly different from unity" but have an economically substantial magnitude.

It is certainly desirable to know how reliable a coefficient estimate is, that is, to know its standard error. But that desirability does not suffice to justify a hypothesis test, which involves measuring the estimate relative to its standard error. Rather, the confidence interval for β_j, constructed from the point estimate b_j and its standard error $\hat{\sigma}_{b_j}$, will be the proper target in most research.

When a null, say, $\beta_j = 1$, is specified, the likely intent is that β_j is *close* to 1, so close that for practical purposes it may be treated *as if it were* 1. But whether 1.1 is "practically the same as" 1.0 is a matter of economics, not of statistics. One cannot resolve the matter by relying on a hypothesis test, because the test statistic $(b_j - 1)/\hat{\sigma}_{b_j}$ measures the estimated coefficient in standard error units, which are not the meaningful units in which to measure the economic parameter $\beta_j - 1$. It may be a good idea to reserve the term "significance" for the statistical concept, adopting "substantial" for the economic concept.

There is a further objection to the common practice of indiscriminately reporting all the "*t*-statistics" for a regression: it encourages rank-ordering of the explanatory variables with respect to their "importance." What does it mean to say that in a multiple regression one explanatory variable is "more important" than another?

A simple example may help to address this question. Suppose that this estimated regression is reported:

$$\hat{y} = 50 + 2x_2 - 1x_3.$$

A naive reader might conclude that x_2 is "more important" than x_3 because its coefficient is larger in magnitude. A more sophisticated reader would recognize that the magnitude of the coefficients can be changed arbitrarily by changing the units in which the variables are measured. So he might ask for the standard errors. Being told that the standard errors for b_2 and b_3 are both 0.5, so their "t-statistics" are 4 and -2, he might conclude that x_2 is "more important" than x_3 because its "t-statistic" is larger in magnitude. But that conclusion is not sensible if in fact the variables are y = weight (in pounds), x_2 = height (in inches), x_3 = exercise (in hours per week), and the regression is to be used by a physician to advise an overweight patient. Would either the physician or the patient be edified to learn that height is "more important" than exercise in explaining variation in weight?

The moral of this example is that statistical measures of "importance" are a diversion from the proper target of the research—estimation of relevant parameters—to the task of "explaining variation" in the dependent variable.

22.6. Using Asymptotics

In the CNR model, provided that $n - k$ is large, there is no need to refer to the t- and F-tables when σ^2 is unknown. Recall the two asymptotic results shown in Section 18.3:

(1) If $u \sim t(n)$, then $u \xrightarrow{D} \mathcal{N}(0, 1)$.
(2) If $v \sim F(m, n)$, then $mv \xrightarrow{D} \chi^2(m)$.

Applied to the CNR model, (1) implies that there is no objection, when $n - k$ is large, to treating

$$\hat{z}_j = u_j = (b_j - \beta_j)/\hat{\sigma}_{b_j}$$

as if it were

$$z_j = (b_j - \beta_j)/\sigma_{b_j}.$$